

WarpStream

Beyond Kafka: Cutting Costs and Complexity with WarpStream and S3

Hidden Cost of Apache Kafka

- Cloud disks are expensive.
- Long retention workloads can be 80% disk cost even at low throughput
- EBS vs instance storage doesn't matter
 - double/triple replication still hugely expensive on both compared to S3

EBS (GP2) \$0.1/GiB pre-replication

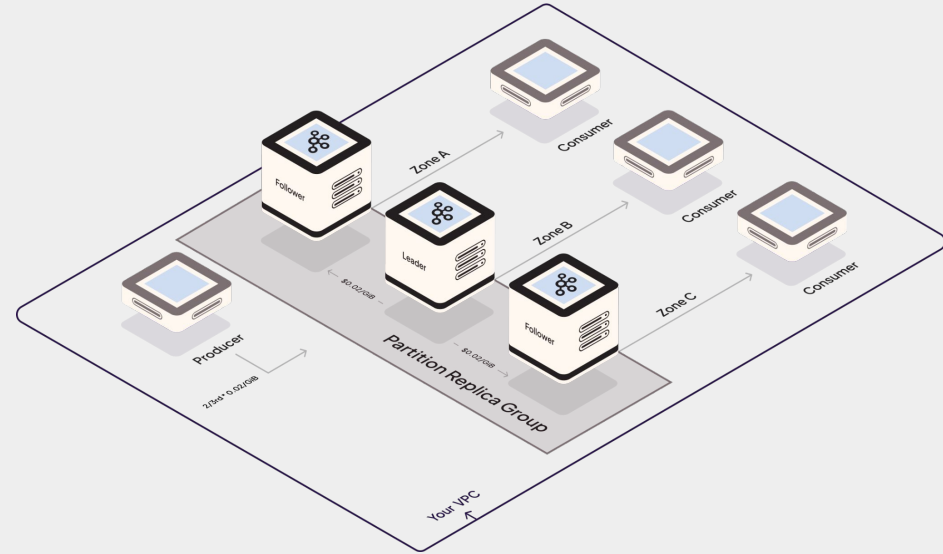
EBS (GP2) \$0.3/GiB post-replication

S3 \$0.02/GiB post-replication

Tiered storage helps, but not enough

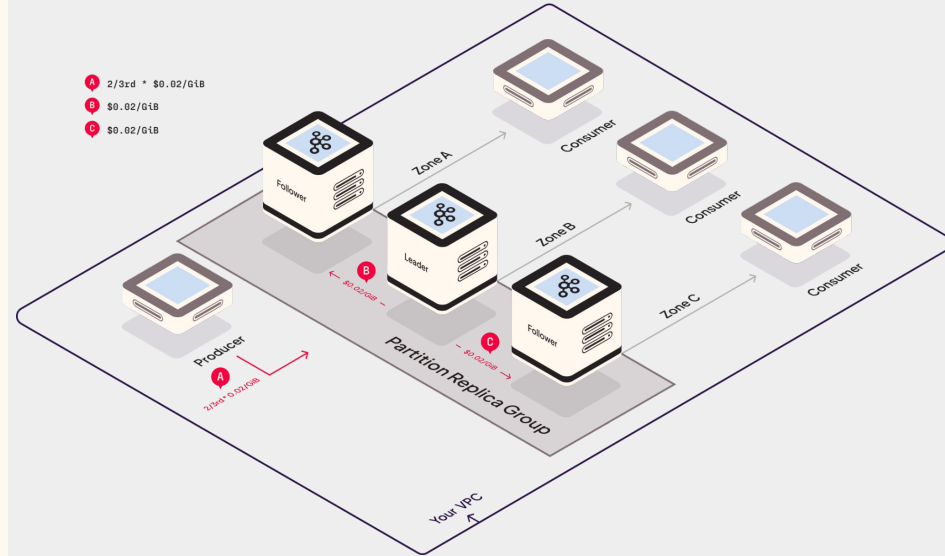
Why drop the disk? Operations

- **Stateful brokers** with attached storage make operations complex, difficult, and inelastic
- Requires consensus, topic-partition leaders, custom operations for scaling in/out and doing node replacements
- Balancing in general always a problem



Why drop the disks? Networking

- 80%+ of TCO for high throughput Apache Kafka clusters can be networking fees
- \$0.053 / compressed GiB transferred in 100% ideal conditions with fetch-from-follower enabled



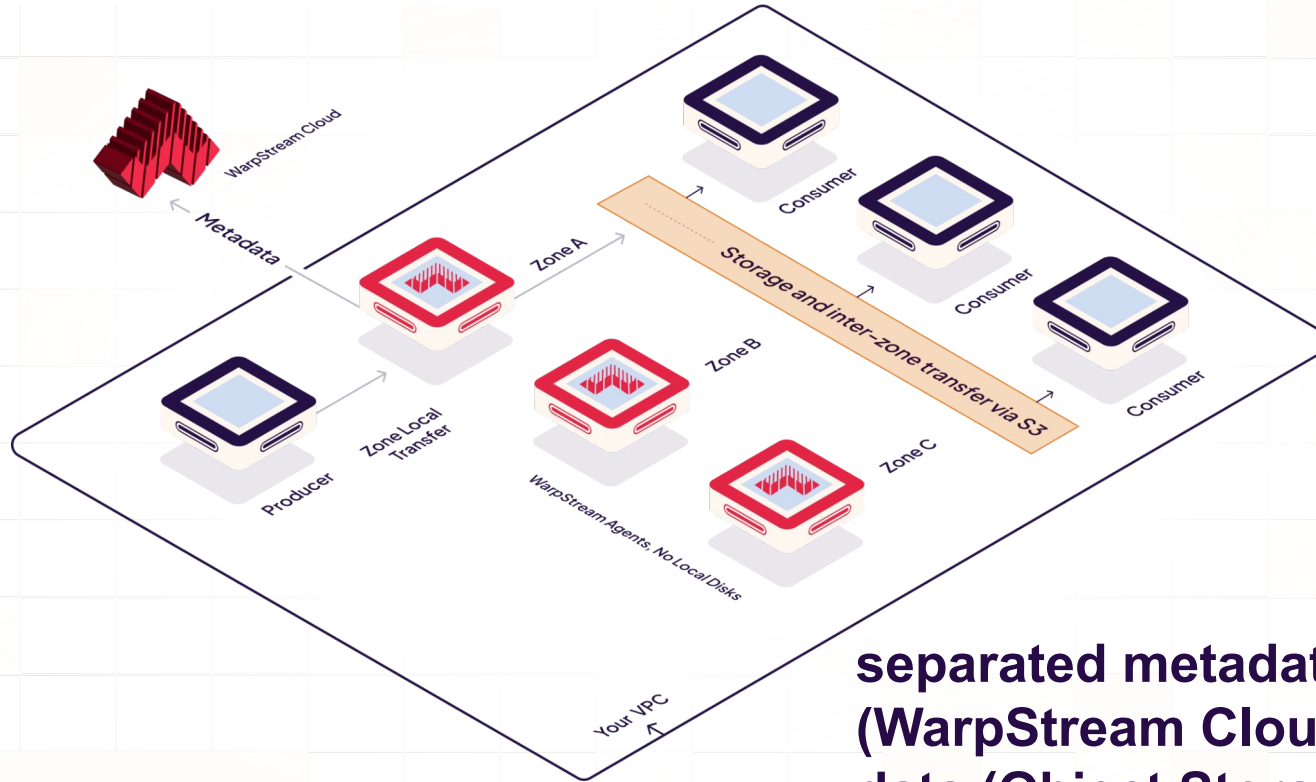
Zero disks would be better

WarpStream's Cloud-Native design



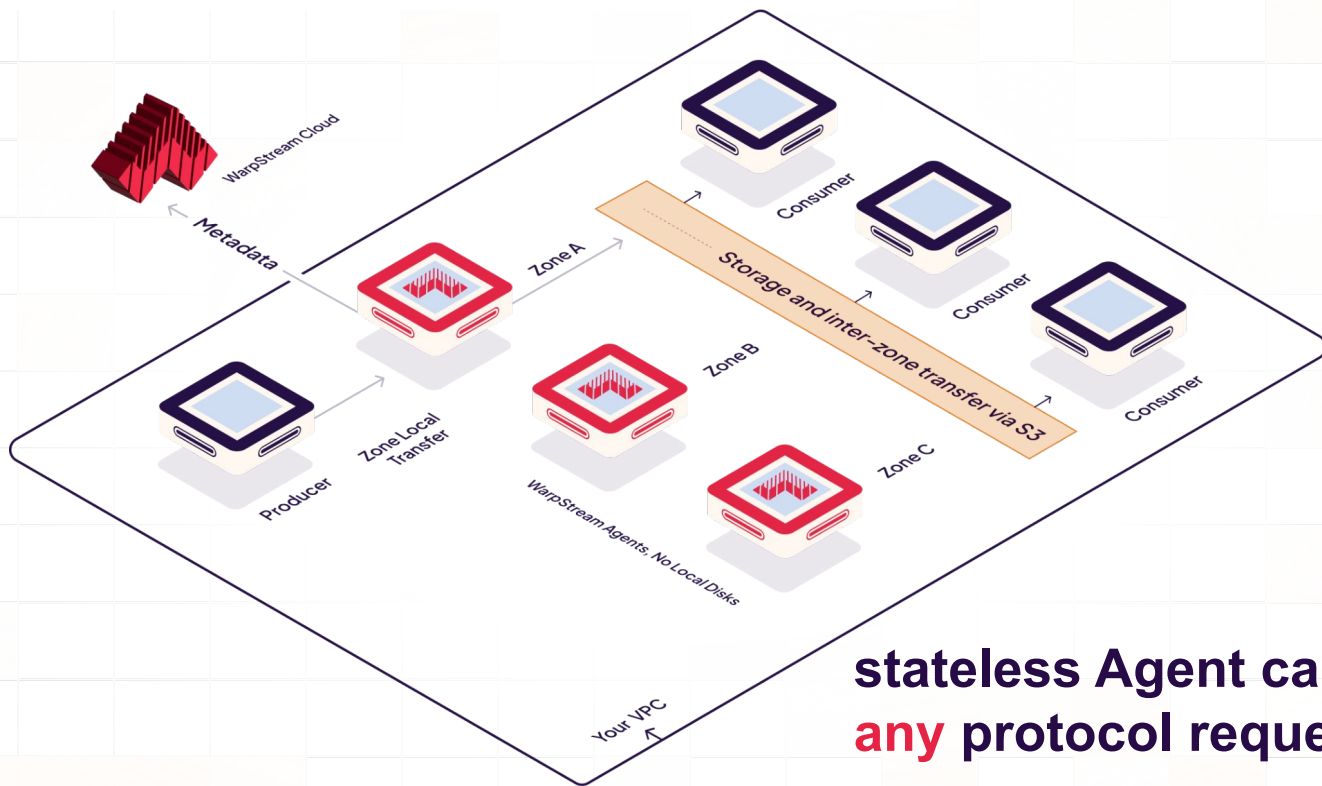
reduces the **cost and complexity** of Apache Kafka

Solving Ease of Operations



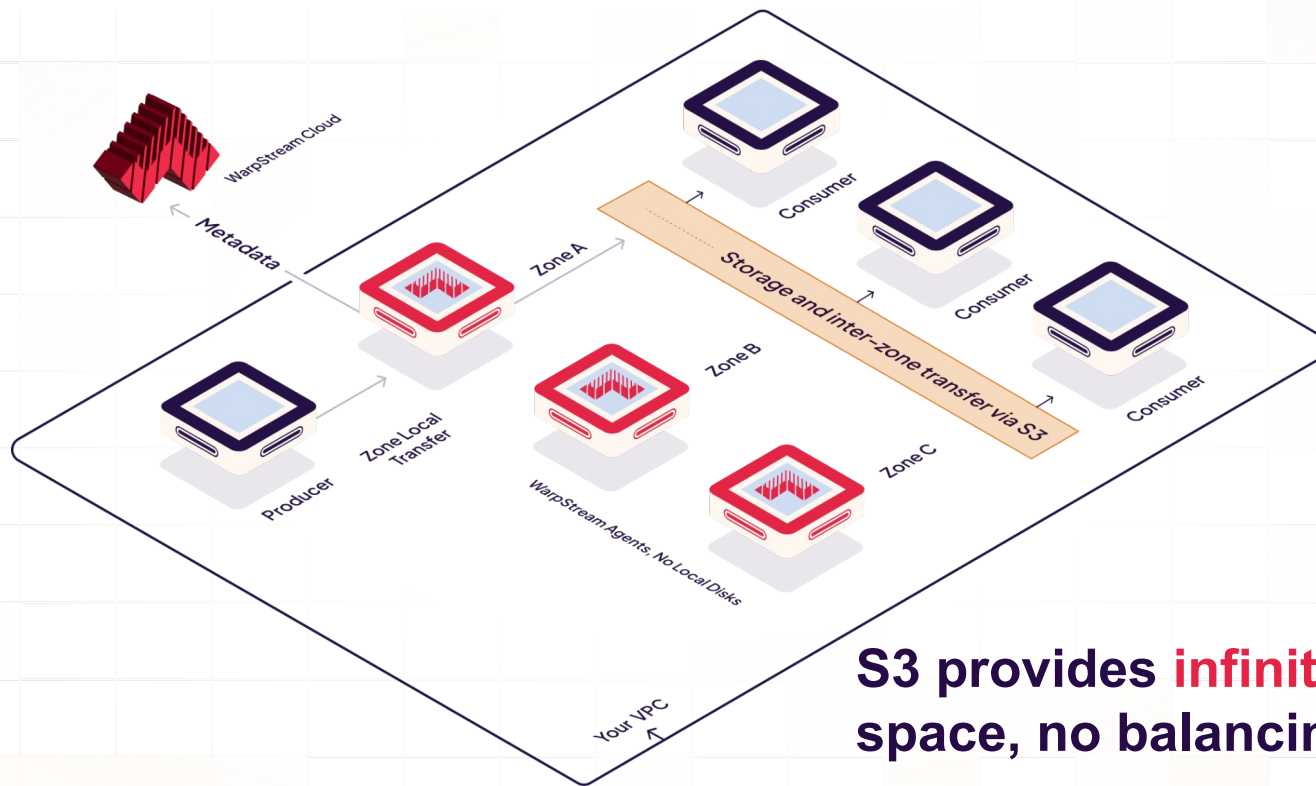
**separated metadata
(WarpStream Cloud) from
data (Object Storage)**

Solving Ease of Operations



stateless Agent can serve
any protocol request

Solving Ease of Operations



S3 provides infinite disk space, no balancing required

How it works

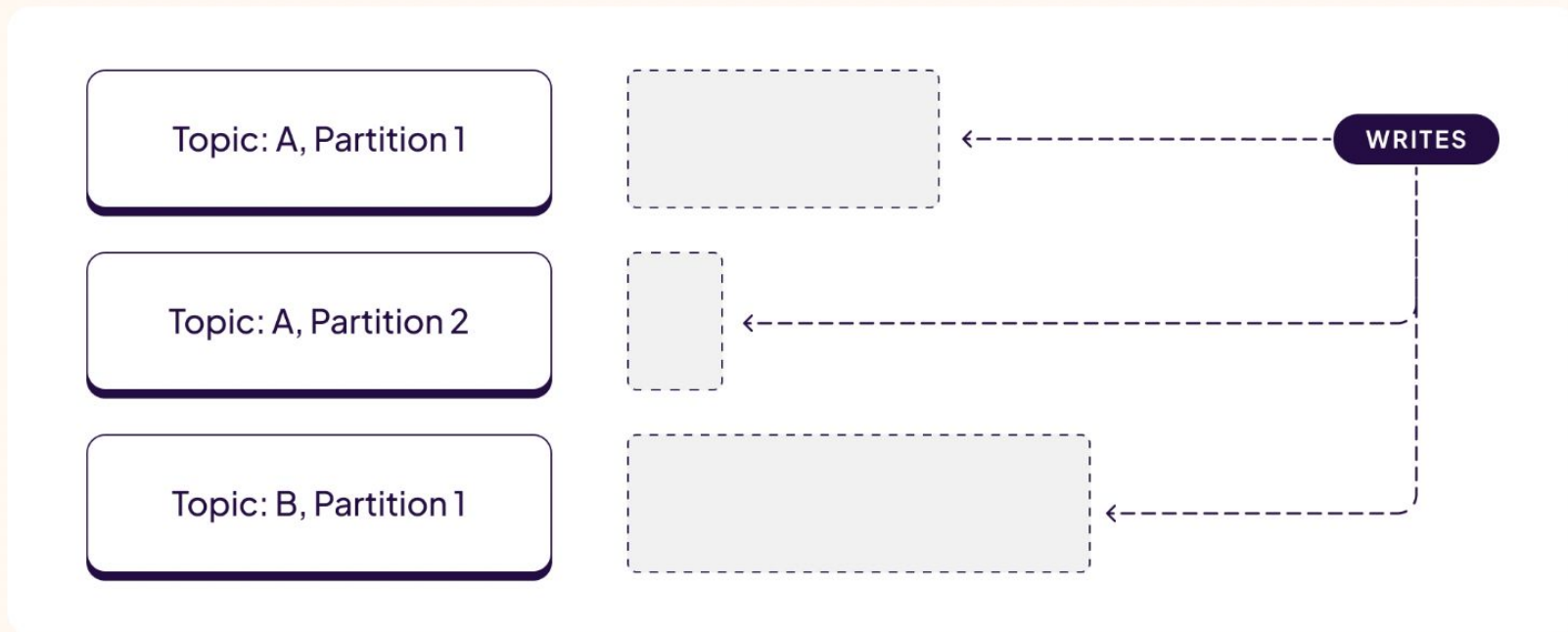
Optimize for cloud unit economics

- Entire storage engine redesigned around minimizing PUT / GET operations
- S3 PUT: \$0.000005
- S3 GET: \$0.000004

Optimize for cloud unit economics

- Entire storage engine redesigned around minimizing PUT / GET operations
- Networking is free, and storage is cheap.
- S3 PUT: \$0.000005
- S3 GET: \$0.0000004
- S3 Storage: \$0.023/GB-mo
- S3 Cross-AZ Networking: **Free**

Step 1: Eliminate topic-partition files



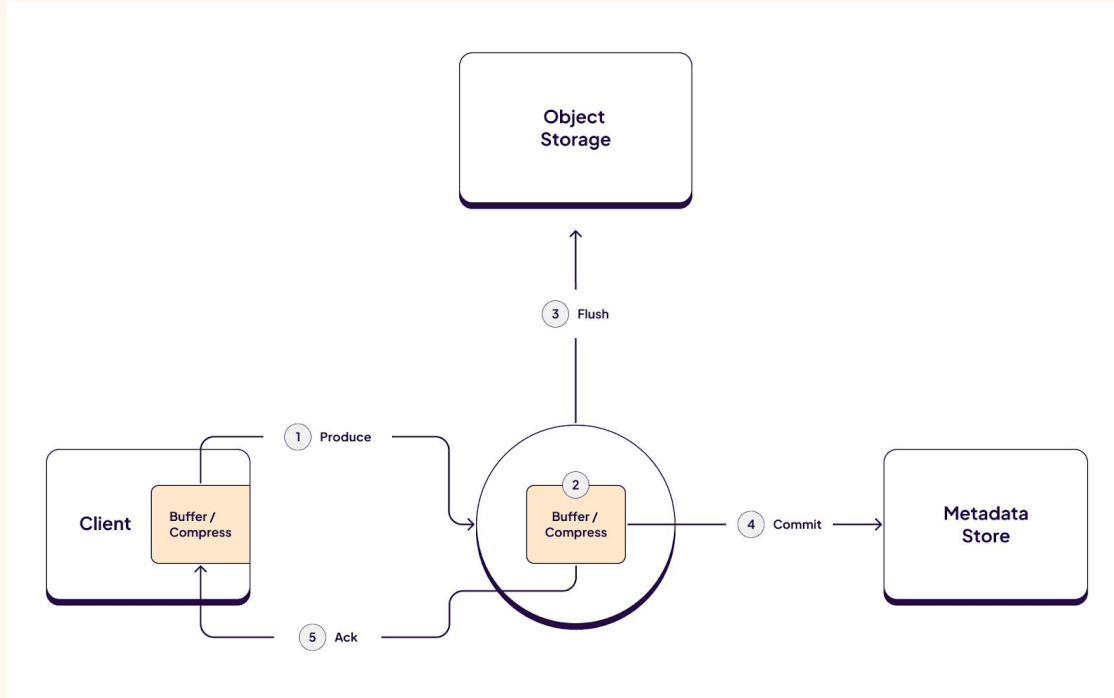
Per topic-partition segment files in Kafka's storage engine.

Step 1: Eliminate topic-partition files

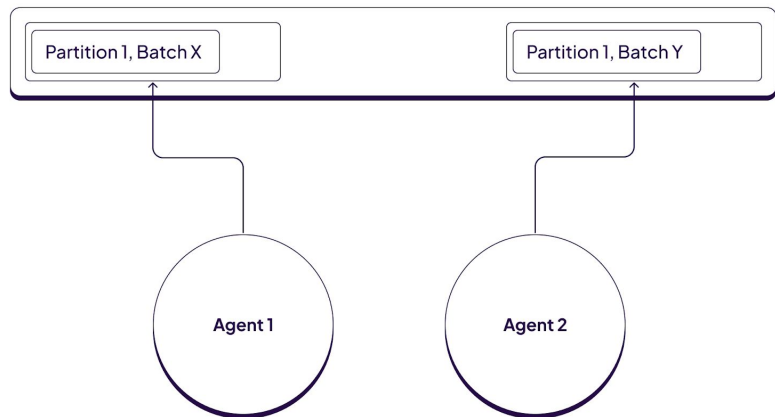


A very high level overview of a typical WarpStream file created by a single Agent during data ingestion. Non sequential partitions indicate that the Agent did not receive any data for some partitions during the elapsed time window.

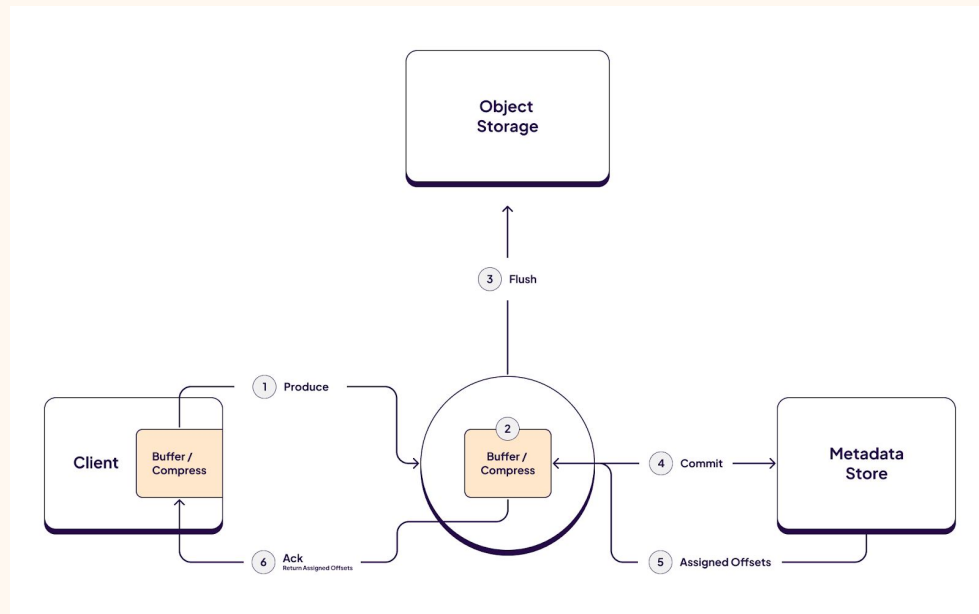
Step 1: Eliminate topic-partition files



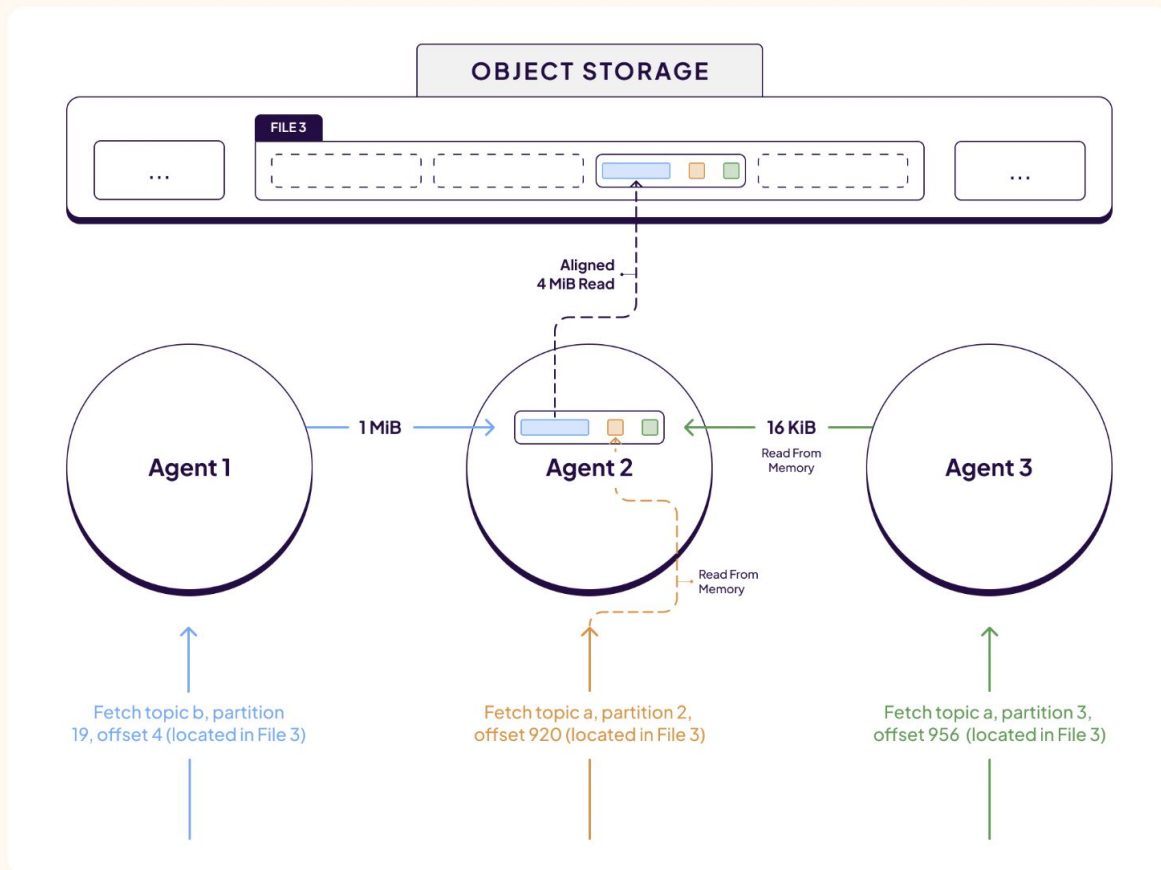
Step 2: Separate data from Metadata



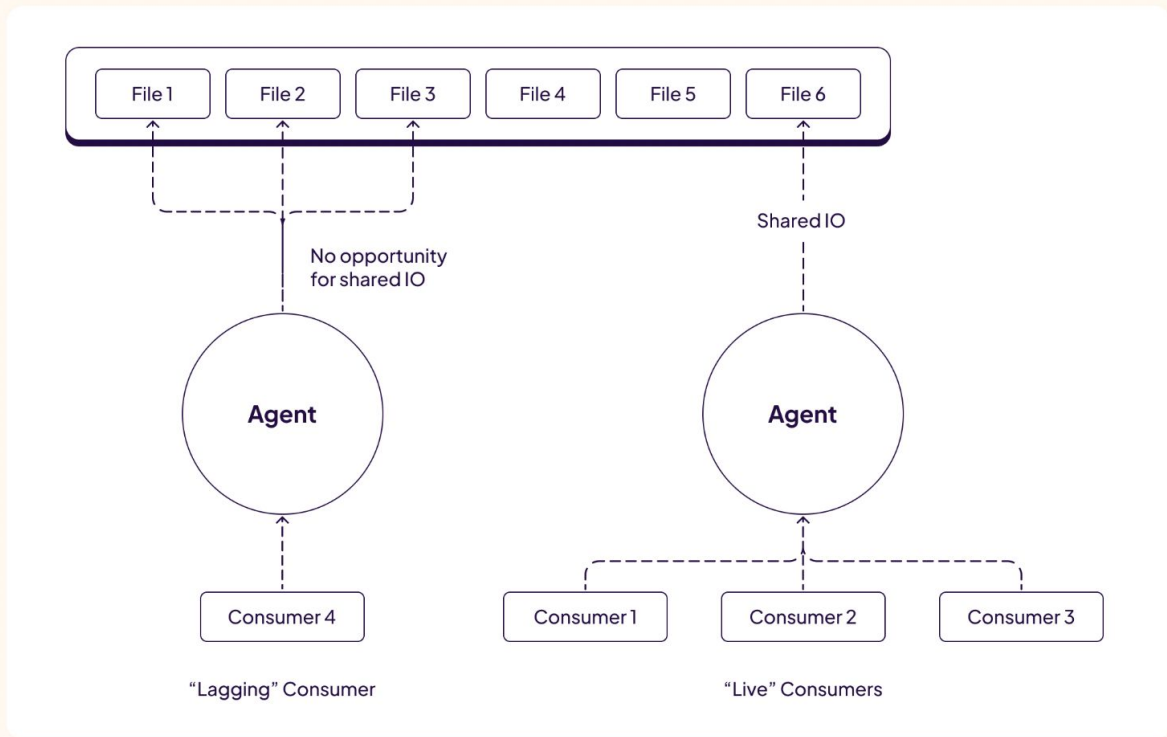
Look ma, no leaders!



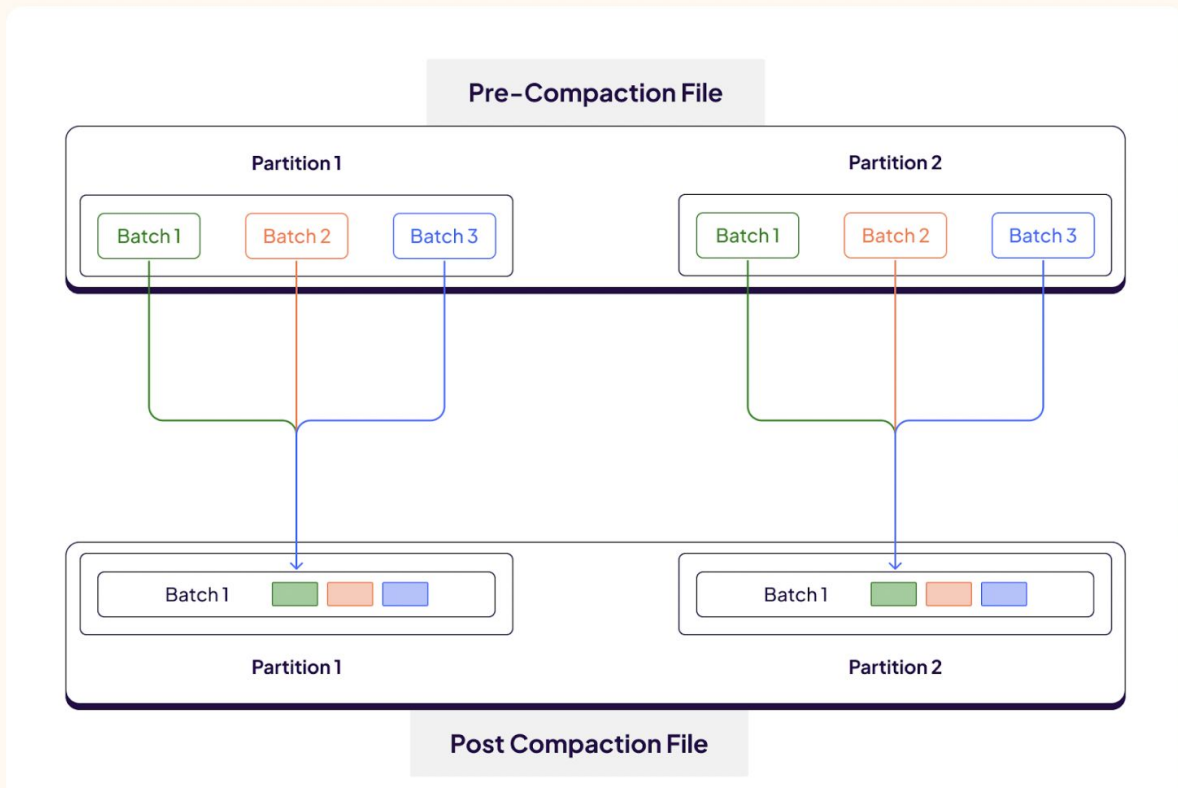
Step 3: Introduce data locality for live reads



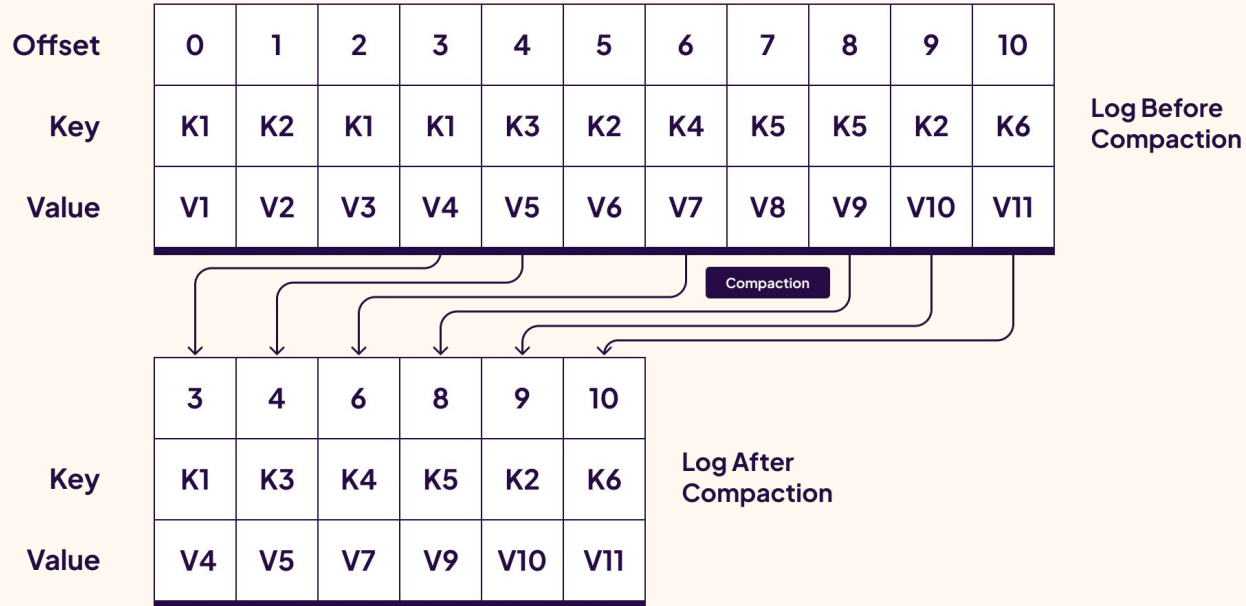
Step 4: introduce data locality for historical reads



Step 4: introduce data locality for historical reads



Hard Mode: Compacted Topics

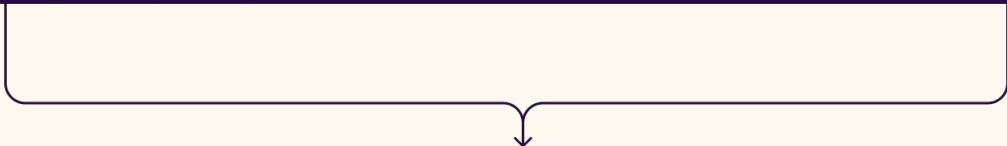


Hard Mode: Compacted Topics

- Tiered Storage in open source Apache Kafka does not support compacted topics
- WarpStream already does compaction internally
- ... how hard could it possibly be?

Hard Mode: Compacted Topics

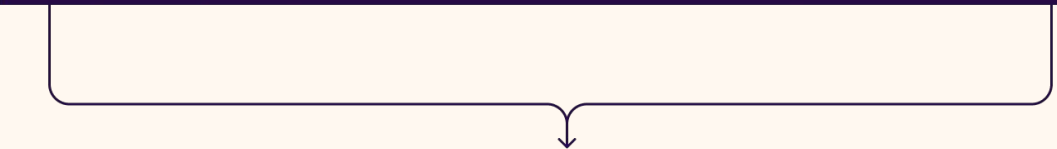
| | | | | | |
|--------|----|----|-----|----------|---------|
| Offset | 3 | 4 | ... | 1000000 | 1000001 |
| Key | K3 | K4 | ... | K1000000 | K3 |



These two records are very far apart in “key space” (almost 10 million unique keys separate them) and will be difficult to deduplicate.

Hard Mode: Compacted Topics

| | | | | | |
|--------|----|----|-----|---------|---------|
| Offset | 3 | 4 | ... | 1000000 | 1000001 |
| Key | K3 | K3 | ... | K3 | K3 |

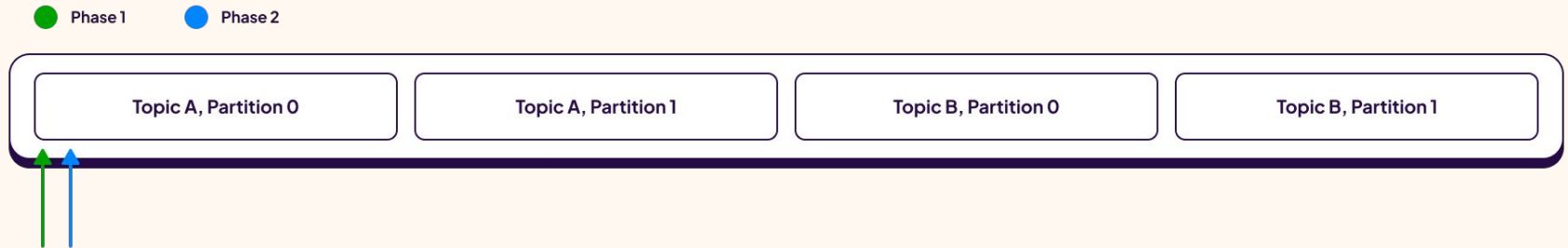


These two records are very far apart in "offset space" (almost 10 million records separate them) but they are very close to each other in "key space" (very few unique keys separate them), so they will be easy to deduplicate.

Hard Mode: Compacted Topics



Hard Mode: Compacted Topics



Hard Mode: Compacted Topics



Hard Mode: Compacted Topics

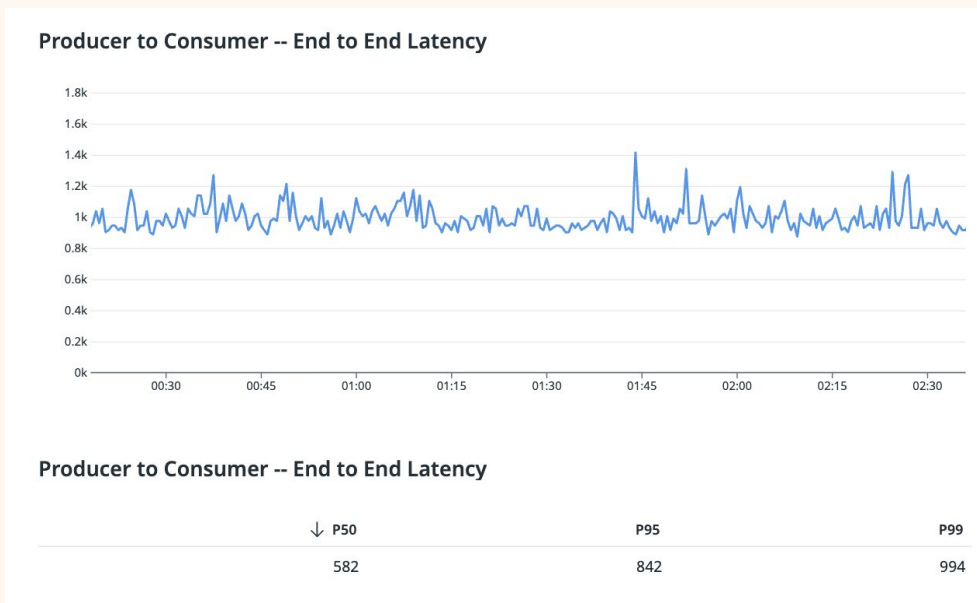


| Deployment Model | Workload Profile | Hardware | Network | Object Storage | Total Costs |
|--------------------------|---|-------------|--------------|----------------|-------------|
| WarpStream | Avg. ingress: 1 GiB/sec Avg. egress: 3 GiB/sec Retention: 1 day Replication Factor = 3 3 Availability Zones | \$223k/year | \$<2k/year | \$61k/year | \$286k/year |
| Self Hosted Apache Kafka | | \$223k/year | \$1.68M/year | \$0 | \$1.9M/year |

WarpStream costs **~85% less** than self-hosted Kafka for high volume workloads

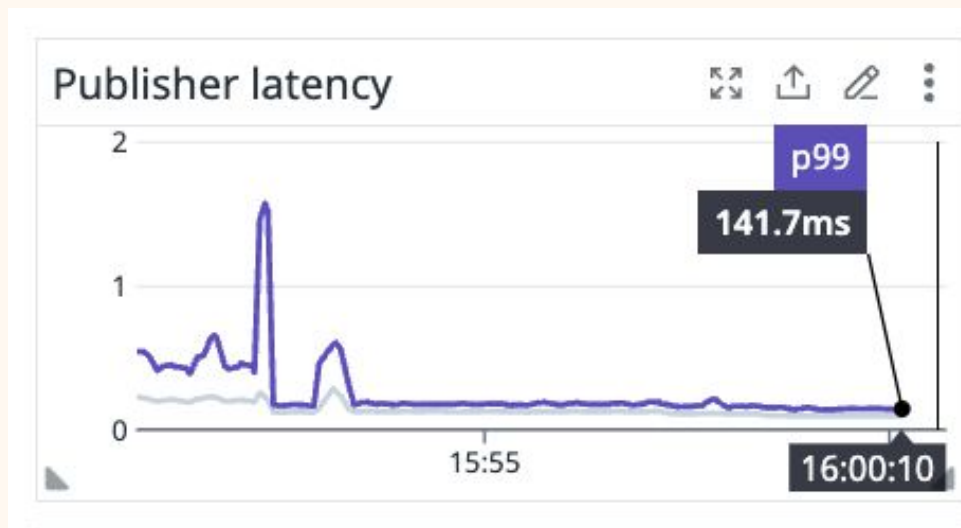
WarpStream is still real time

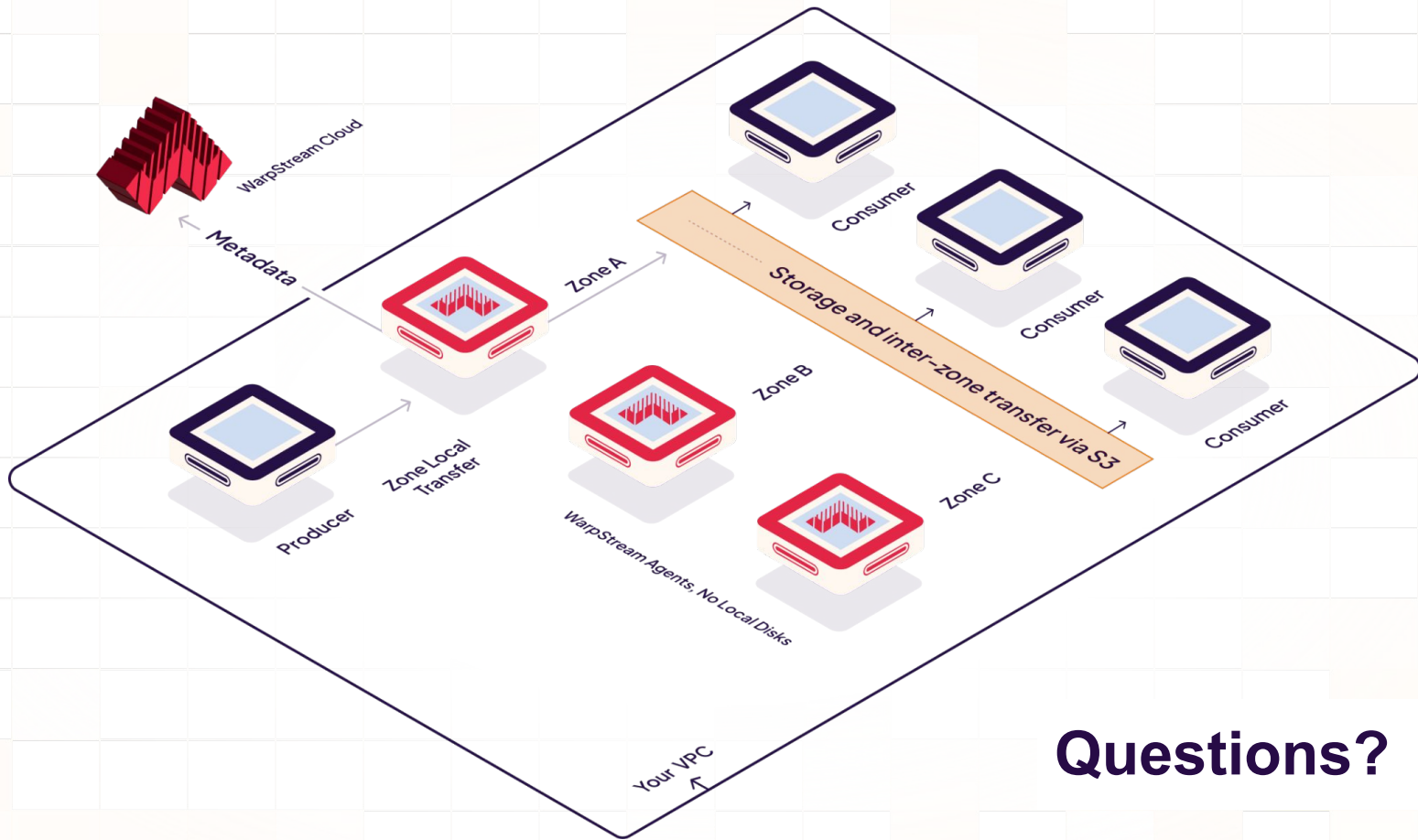
- P99 producer latency of ~400ms
- Producer to Consumer – End to End Latency <1.5s



WarpStream supports S3 Express One Zone

- P99 producer latency as low as 150ms
- Uses a majority quorum of 3 buckets to provide regional high-availability
- Data is moved to S3 Standard asynchronously to reduce storage costs





Questions?

Live Demo

showing WarpStream in action at high throughput



WarpStream